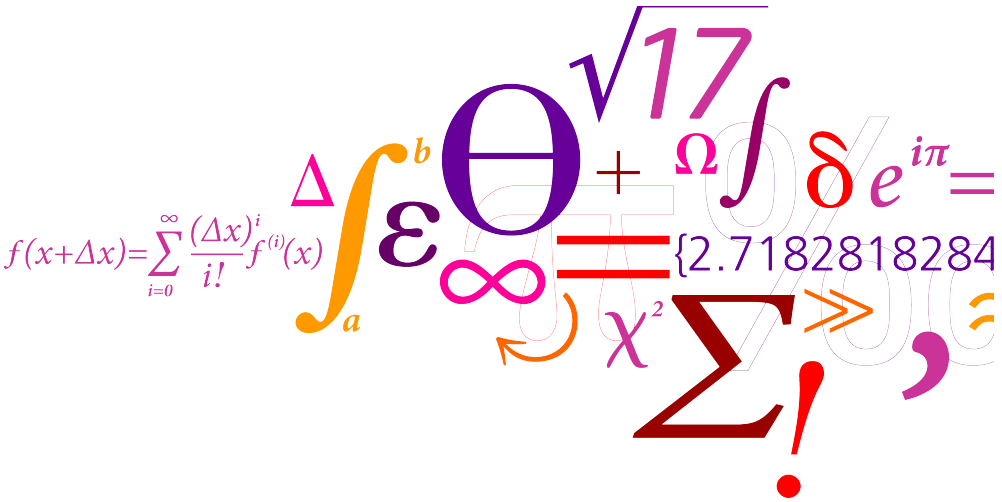
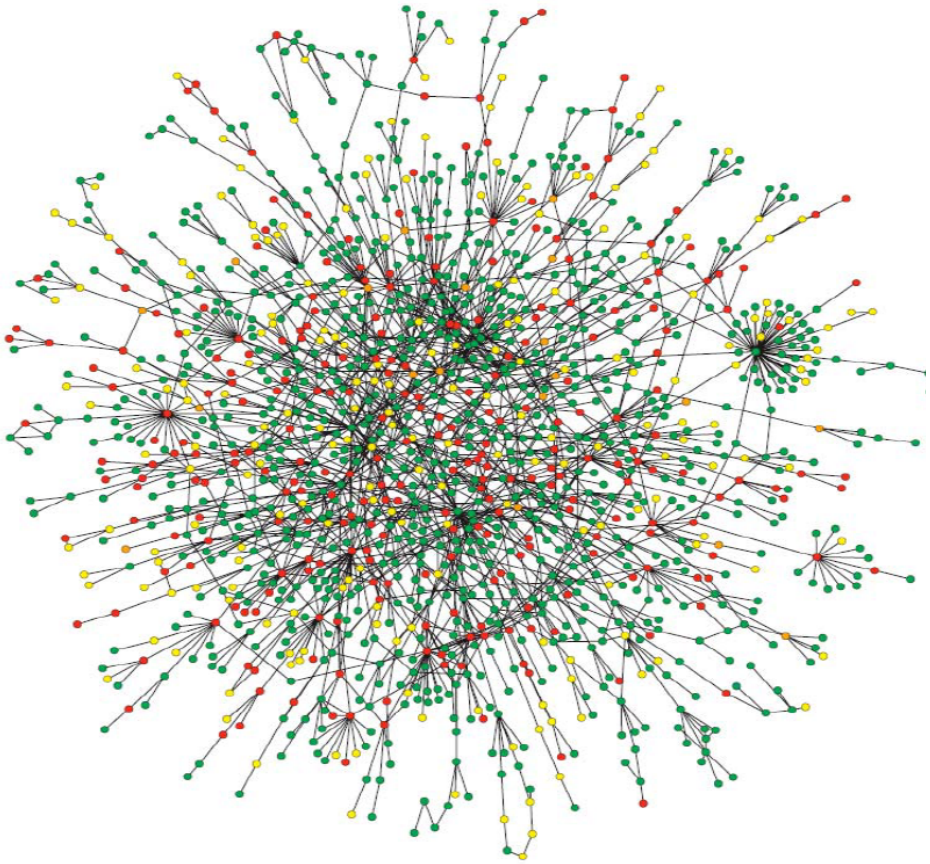
Network Properties

Greg Slodkowicz
Regulatory Genomics group, CBS



Center for Biological Sequence Analysis
Department of Systems Biology

Interaction networks in molecular biology

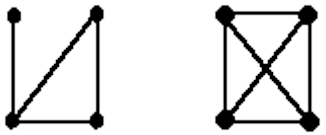


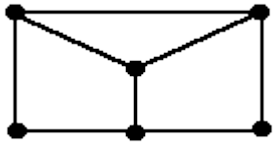
Barabasi & Oltvai, Nature Reviews, 2004

- Protein-protein interactions
- Protein-DNA interactions
- Genetic interactions
- Metabolic reactions
- Co-expression interactions
- Text mining interactions
- Association networks

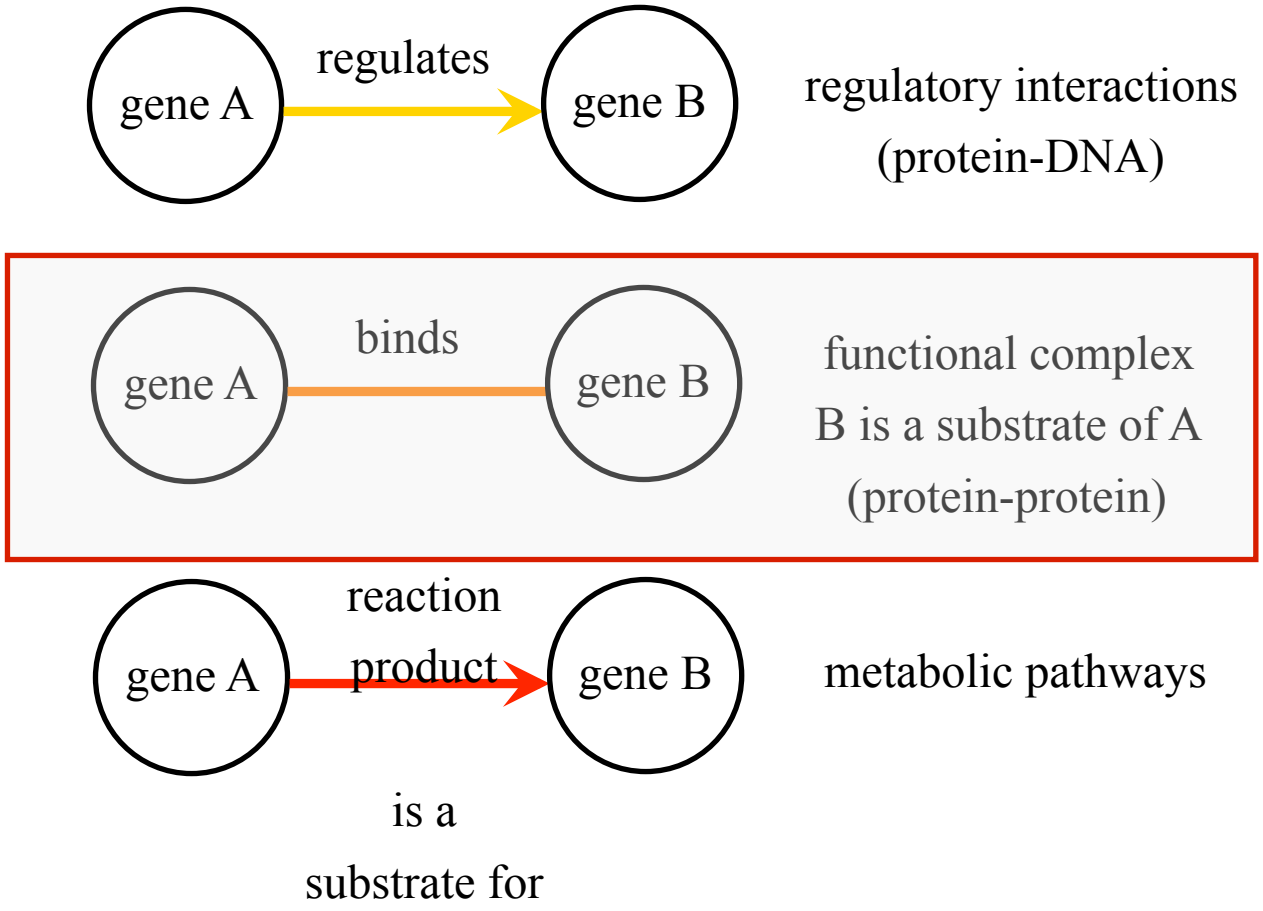
Graphs

- Graph $G=(V, E)$ is a set of vertices V and edges E
- A subgraph G' of G is $V' \subset V$ and $E' \subset E$
- Graph properties:
  - Connectivity (node degree, paths)
  - Cyclic vs. acyclic
  - Directed vs. undirected

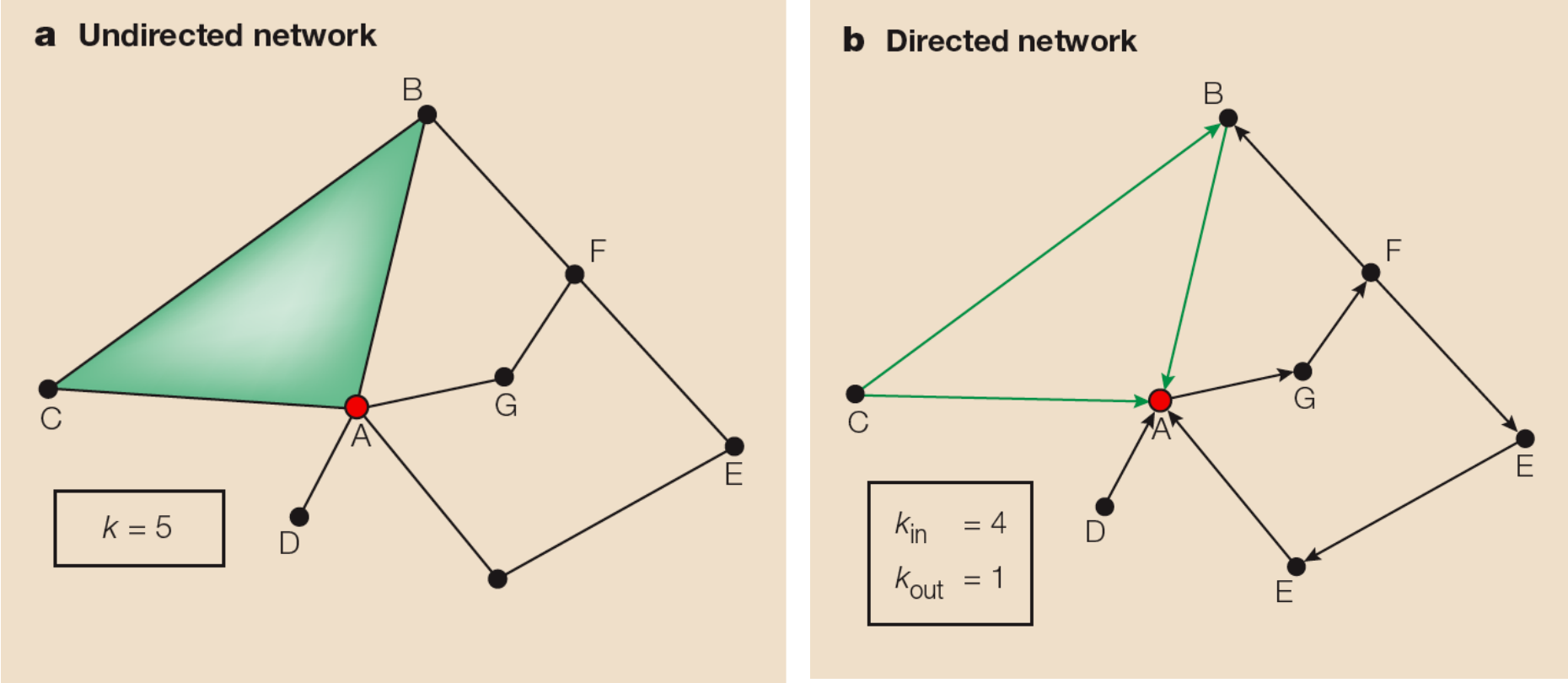




Protein network representations

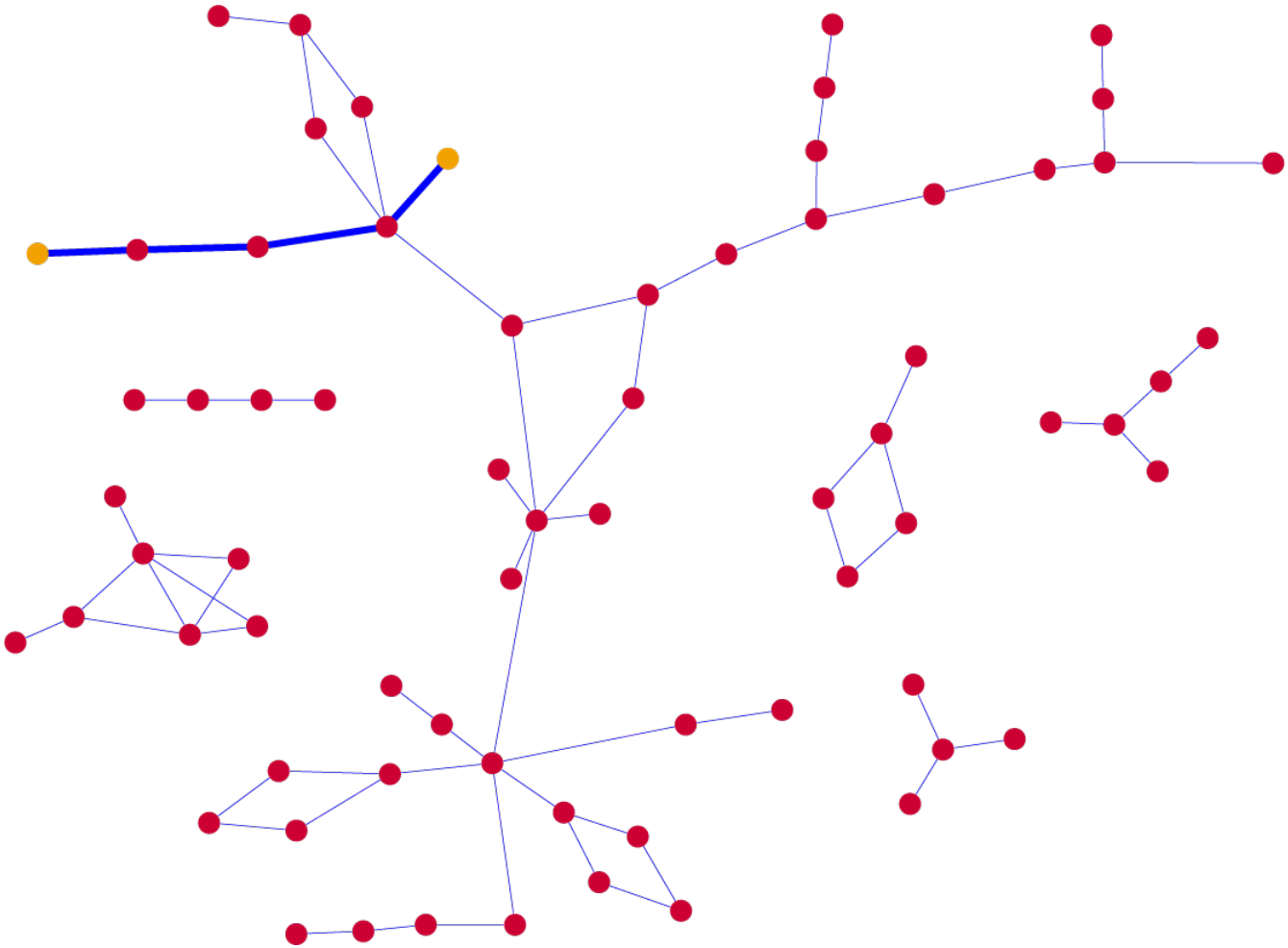


Degree or connectivity

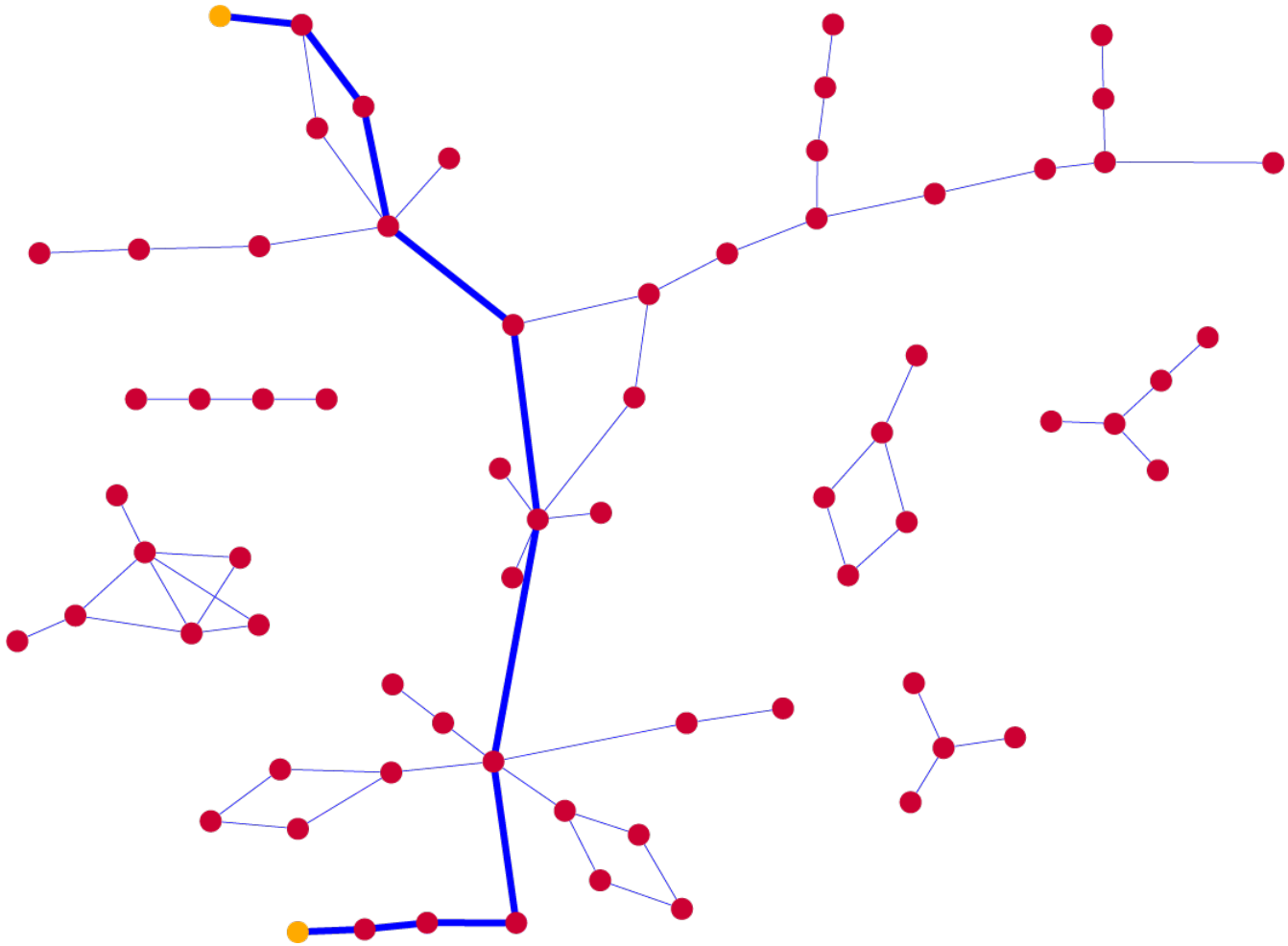


Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet.* 2004 Feb;5(2):101-13

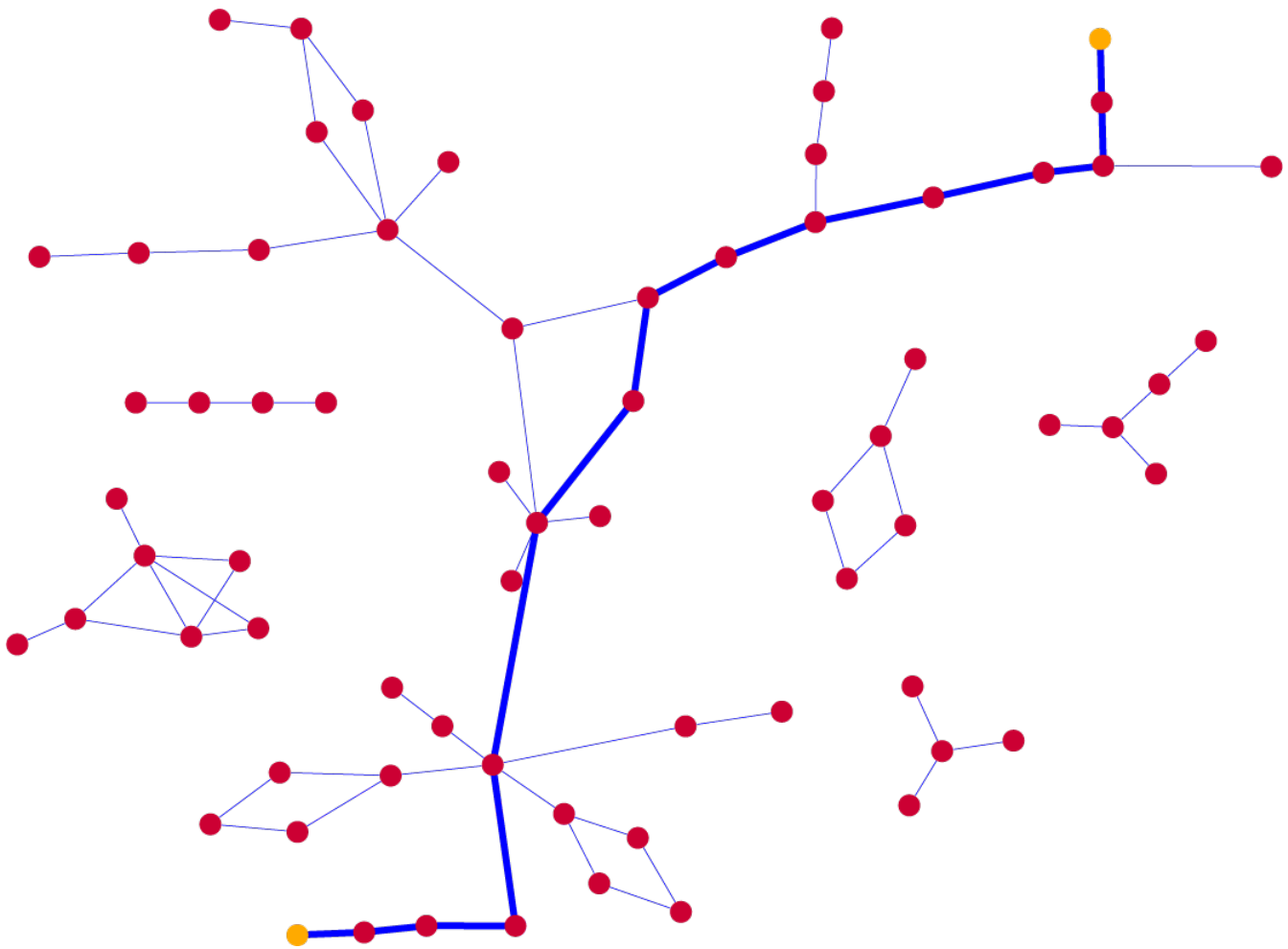
Shortest-Path between nodes



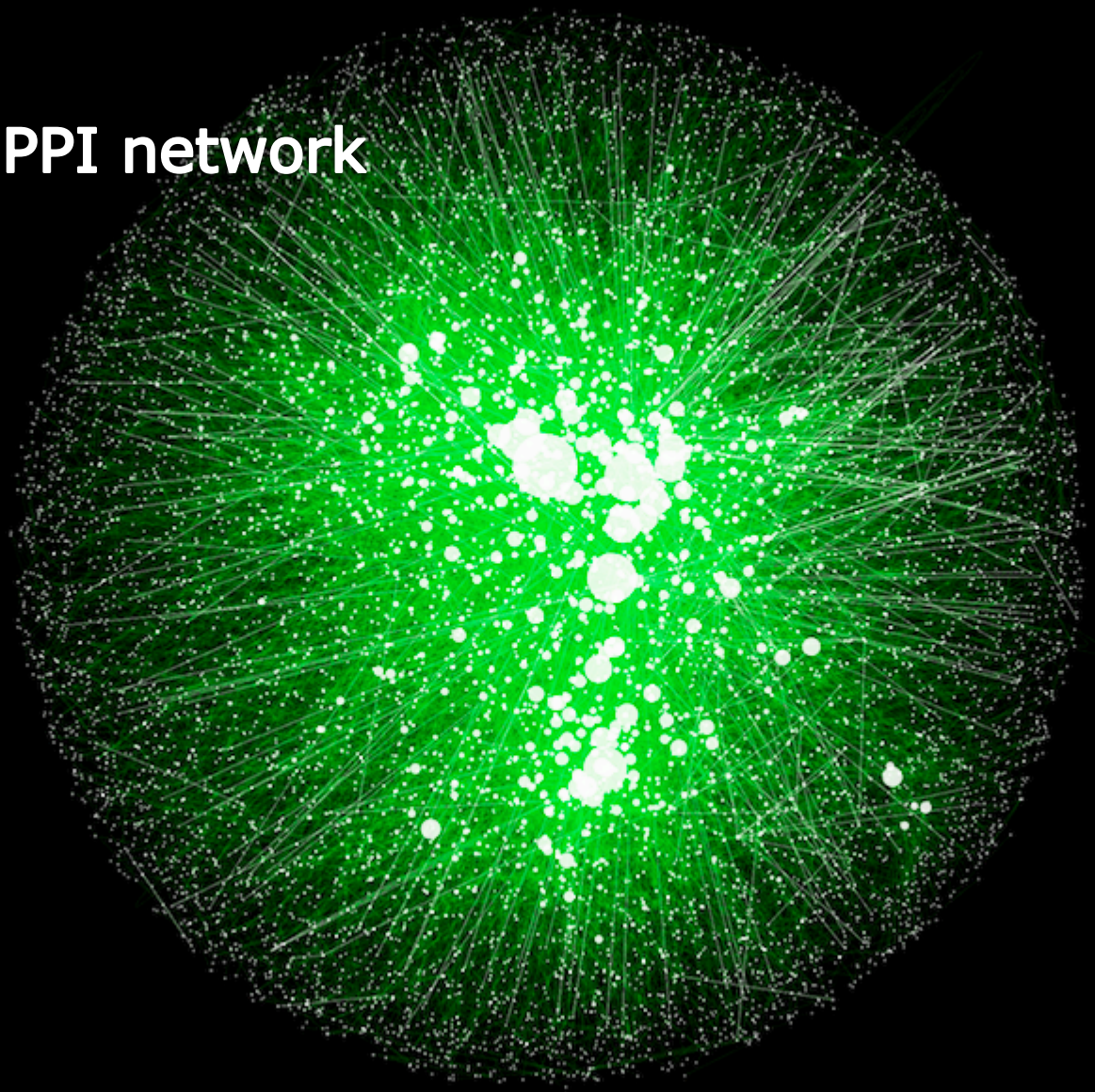
Shortest-Path between nodes



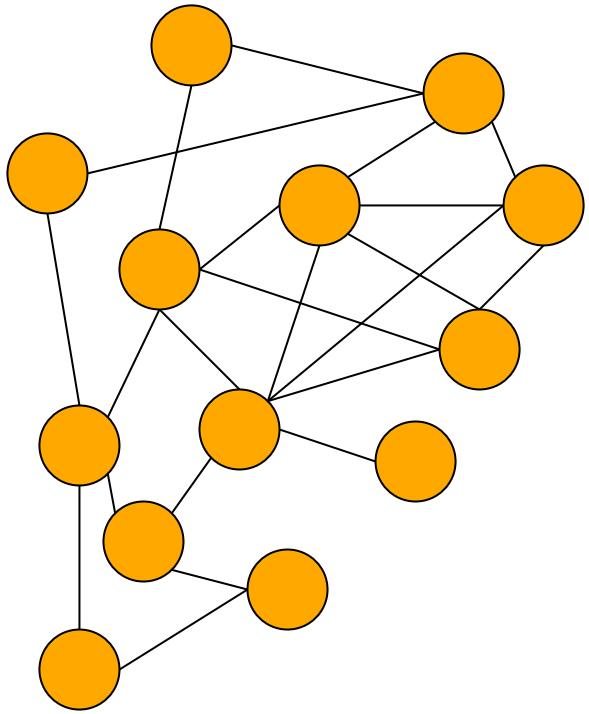
Longest Shortest-Path



Human PPI network



The MCODE algorithm



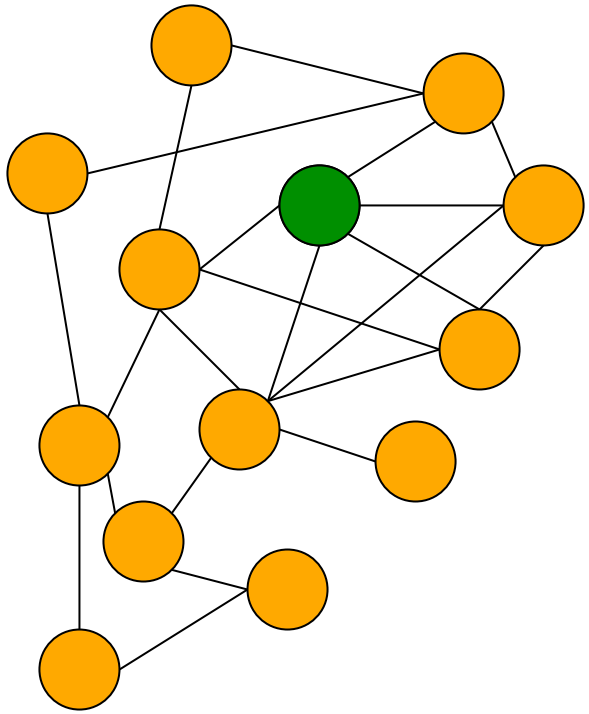
Molecular Complex Detection

MCODE

The three steps of MCODE

1. **Vertex weighting**
2. **Complex prediction**
3. **Post-processing**

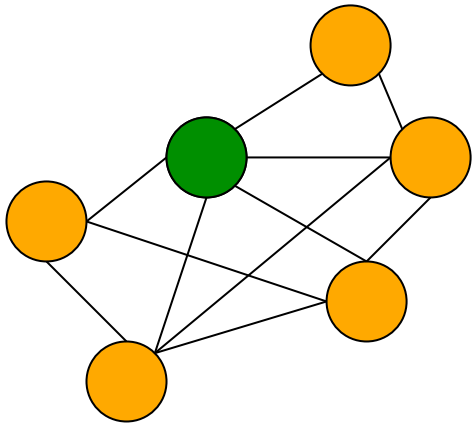
Vertex (nodes) weighting



Vertex weighting

1. **Find neighbors**

Vertex (nodes) weighting



Vertex weighting

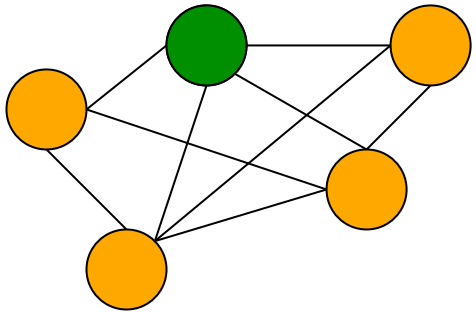
1. **Find neighbors**
2. **Get highest *k*-core graph**

***K*-core graph:**

A graph of minimal degree *k, i.e.*

All nodes must have at least *k* connections

Vertex (nodes) weighting



Vertex weighting

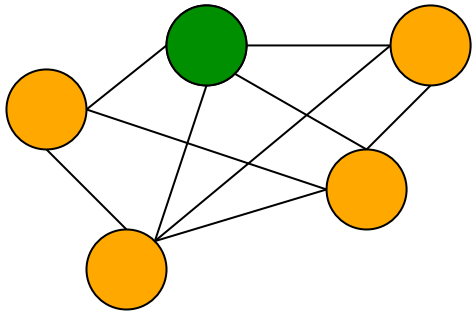
1. **Find neighbors**
2. **Get highest *k*-core graph**
3. **Calculate *density* of k-core graph**

Density:

Number of observed edges, E, divided by the total number of possible edges, E_{max}

$E_{max} = V (V-1)/2$ (networks without loops)

Vertex (nodes) weighting



Vertex weighting

1. **Find neighbors**
2. **Get highest *k*-core graph**
3. **Calculate *density* of k-core graph**
4. **Calculate vertex (node) weight:**

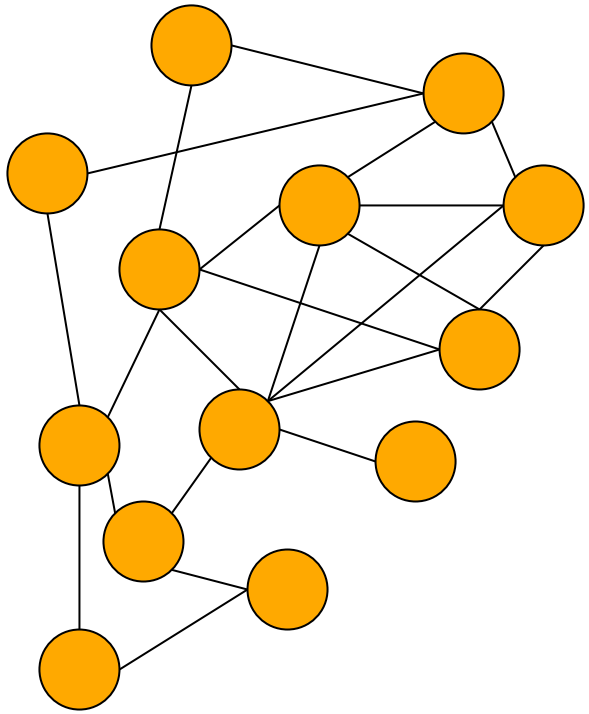
   Density ***** k_{max}

Density:

Number of observed edges, E, divided by the total number of possible edges, E_{max}

$E_{max} = V (V-1)/2$ **(networks without loops)**

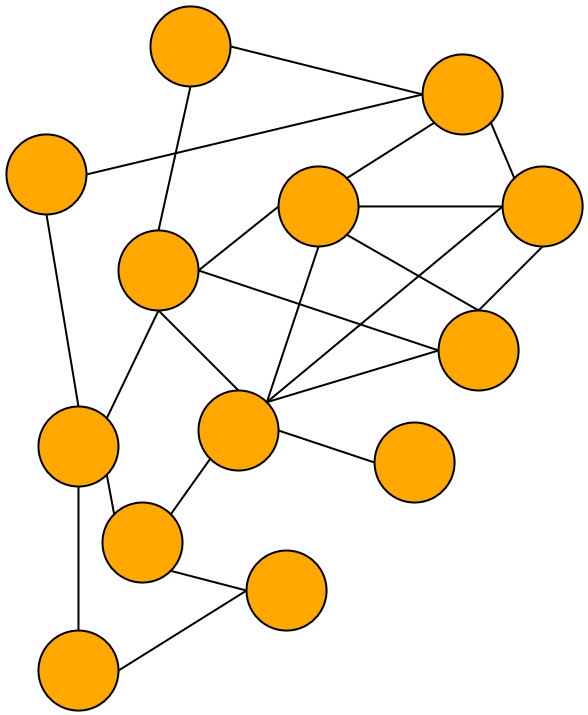
Molecular complex prediction



Complex prediction

1. Seed complex by nodes with highest weight
2. Include neighbors if the vertex weight is above threshold (VWP)
3. Repeat step 2 until no more nodes can be included

To read more about MCODE



BMC Bioinformatics

BioMed Central

Methodology article

Open Access

An automated method for finding molecular complexes in large protein interaction networks

Gary D Bader[1,2] and Christopher WV Hogue*[1]

Address: [1]Samuel Lunenfeld Research Institute, Mt. Sinai Hospital, Toronto ON Canada M5G 1X5, Dept. of Biochemistry, University of Toronto, Toronto ON Canada M5S 1A8 and [2]Current address: Memorial Sloan-Kettering Cancer Center 1275 York Avenue, Box 460, New York, NY, 10021, USA

Email: Gary D Bader - gary.bader@utoronto.ca; Christopher WV Hogue* - hogue@mshri.on.ca

* Corresponding author

Published: 13 January 2003

BMC Bioinformatics 2003, 4:2

Received: 4 September 2002

Accepted: 13 January 2003

This article is available from: http://www.biomedcentral.com/1471-2105/4/2

Abstract

Background: Recent advances in proteomics technologies such as two-hybrid, phage display and mass spectrometry have enabled us to create a detailed map of biomolecular interaction networks. Initial mapping efforts have already produced a wealth of data. As the size of the interaction set increases, databases and computational methods will be required to store, visualize and analyze the information in order to effectively aid in knowledge discovery.

Results: This paper describes a novel graph theoretic clustering algorithm, "Molecular Complex Detection" (MCODE), that detects densely connected regions in large protein-protein interaction networks that may represent molecular complexes. The method is based on vertex weighting by local neighborhood density and outward traversal from a locally dense seed protein to isolate the dense regions according to given parameters. The algorithm has the advantage over other graph clustering methods of having a directed mode that allows fine-tuning of clusters of interest without considering the rest of the network and allows examination of cluster interconnectivity, which is relevant for protein networks. Protein interaction and complex information from the yeast *Saccharomyces cerevisiae* was used for evaluation.

Conclusion: Dense regions of protein interaction networks can be found, based solely on connectivity data, many of which correspond to known protein complexes. The algorithm is not affected by a known high rate of false positives in data from high-throughput interaction techniques. The program is available from ftp://ftp.mshri.on.ca/pub/BIND/Tools/MCODE.

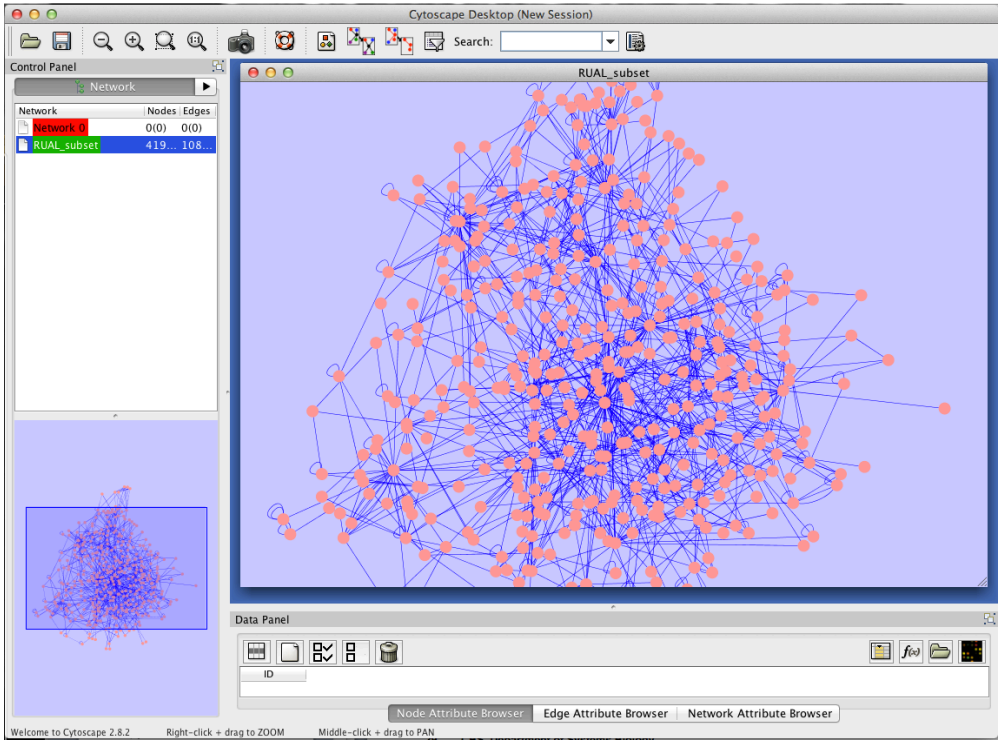
Background

Recent papers published in *Science* and *Nature* among others describe large-scale proteomics experiments that have generated large data sets of protein-protein interactions and molecular complexes [1–7]. Protein structure [8] and gene expression data [9] is also accumulating at a rapid rate. Bioinformatics systems for storage, management, visualization and analysis of this new wealth of data must keep pace. We previously published a simple graph theory method that identified a functional protein complex around the yeast protein Las17 that is involved in actin cytoskeleton rearrangement [10]. Here we extend the method to better apply it to the accumulating information in protein networks.

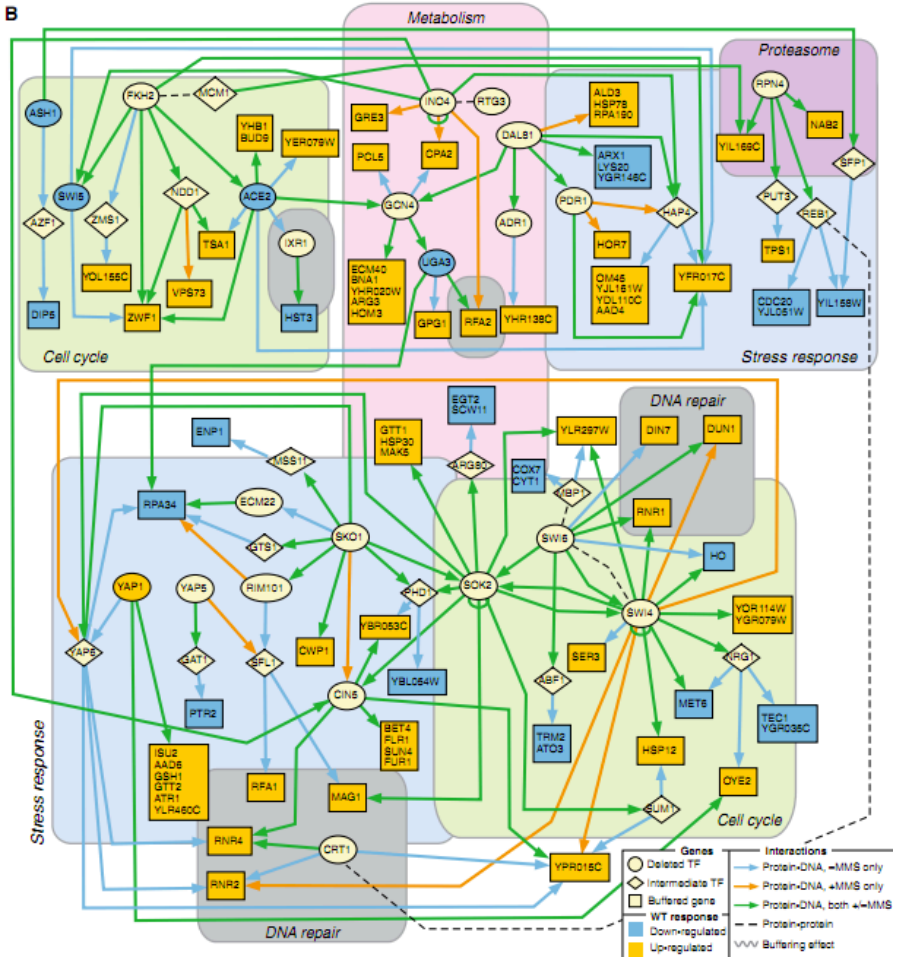
Currently, most proteomics data is available for the model organism *Saccharomyces cerevisiae*, by virtue of the availability of a defined and relatively stable proteome, full

Cytoscape

- Network visualisation and analysis
- Free, open source software
- Popular in bioinformatics/systems biology
- Used in other fields



Cytoscape example



Workman et al. Science 2005

And now on to the Cytoscape exercise

NOTE: We skip **Q5.**

Credit note: most of these slides were created by Prof. Christopher Workman, for the course 27041 Systems Biology